

Abstract

One exemplary embodiment of a scalable clustering algorithm accesses a database of records having attributes or data fields of both enumerated discrete and ordered values and brings a portion of the data records into a rapid access memory. A cluster model for the data includes a table of probabilities for the
5 enumerated, discrete data fields of the data records. The cluster model for data fields that are ordered comprises a mean and spread of the cluster. The cluster model is updated from the database records brought into the rapid access memory. At least some of the database records in the rapid access memory are summarized and stored within the rapid access memory. A criteria is then evaluated to
10 determine if further data should be accessed from the database to further cluster data records in the database. Based on the evaluating step, additional database records in the database are accessed and brought into the rapid access memory for further updating of the cluster model.